Active Shape Models from a Level Set Perspective

Mikaël Rousson, Nikolaos Paragios, Rachid Deriche

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Active Shape Models from a Level Set Perspective

Mikaël Rousson — Nikos Paragios — Rachid Deriche

N° 4984

October 2003

THÈME 3

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--4984--FR+ENG

Active Shape Models from a Level Set Perspective

Mikaël Rousson* , Nikos Paragios † , Rachid Deriche *

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet Odyssee

Rapport de recherche n° 4984 — October 2003 — 18 pages

Abstract: Active shape models [5] is a popular technique to object extraction. To this end, one models the geometric form of the object of interest. Object extraction is then equivalent with seeking a linear/non-linear transformation that projects/deforms this geometric form to an image region with the desired visual properties. The definition of the image term is the most challenging component of such an approach. Level set methods [14] is a powerful optimization framework, that can be used to recover objects of interest by the propagation of curves. They can support complex topologies, considered in higher dimensions, are implicit, intrinsic and parameter free. Furthermore, one can introduce various image terms when seeking an object of particular form/visual properties. In this paper we re-visit active shape models and introduce a level set variant of them. Such an approach can account for prior shape knowledge quite efficiently as well as use data/image terms of various form and complexity while being able to deal with important local deformations and changes of topology. Promising experimental results demonstrate the potential of our approach.

Key-words: Active Shape Models, Level Set Theory, Shape Prior, Segmentation, Object Extraction, Tracking.

* Projet Odyssée, INRIA Sophia-Antipolis, 2004, route des Lucioles, BP 93, 06902 Sophia-Antipolis, France, `{Mikael.Rousson,Rachid.Deriche}@sophia.inria.fr`, `www-sop.inria.fr/odyssee/presentation/index.en.html`

† Project Manager, Siemens Corporate Research, 755 College Road East, Princeton, NJ, 08540, USA, `Nikos.Paragios@scr.siemens.com`, `http://home.comcast.net/~aparagio`

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : +33 4 92 38 77 77 —Télécopie : +33 4 92 38 77 65

Modèles Actifs de Formes par ensembles de niveaux

Résumé : La technique d'extraction d'objets basée sur les "Modèles Actifs de Formes" [5] est très populaire. Elle consiste à modéliser la forme de l'objet dans une première phase, puis à extraire l'objet via la recherche d'une transformation, linéaire ou pas, qui évolue cette forme géométrique afin qu'elle intégre au mieux tout un ensemble d'attributs image. La définition du terme d'attache aux données est fondamentale dans une telle approche. La méthodologie à base d'ensembles de niveaux [14] fournit un cadre d'optimisation efficace, permettant de détecter les objets d'intérets par évolution de courbes. Cette représentation fournit de nombreux avantages par rapport aux approches classiques: elle peut représenter des formes de topologie complexe, être utilisée pour des dimensions plus élevées, est implicite, intrinsèque et n'ajoute aucun paramètre. De plus, divers termes d'attache aux données peuvent être introduits pour détecter un objet ayant une forme ou des caractéristiques visuelles spécifiques.
Dans ce rapport, nous revisitons le "modèle actif de formes" et nous introduisons une variante basée sur les ensembles de niveaux. Cette approche peut intégrer de manière efficace, aussi bien des à priori sur la forme des objets que des termes d'attache aux données de complexité et de type variés, tout en permettant des déformations locales importantes et des changements de topologie. Des résultats expérimentaux prometteurs illustrent les grandes potentialités de cette nouvelle approche.

Mots-clés : Modèles Actifs de Formes, ensemble de niveaux, à priori sur la forme, segmentation, extraction d'objets, suivi d'objets.

1 Introduction

Object extraction from images is a quite popular component of image understanding. Segmentation, tracking, reconstruction [7] and 3D modeling are some application domains that could benefit from such a module. In this paper, we aim to address this application with objective to recover a structure of particular geometric form from an image.
B-splines deformable models as well as point distribution models are mathematical formulations introduced to the snake framework [8] to account for shape consistency. Active shape [5] and appearance [3] models were a major breakthrough in object extraction and image segmentation. Such a framework consists of two stages; (i) the modeling and (ii) the segmentation phase.
During modeling the objective is to recover a compact representation for the geometric form of the structure of interest. Using a set of registered training examples, one can either represent prior knowledge using simple or more complicated density functions. Gaussian distribution [5], mixture models [4] or non-parametric function [6] were considered in the past.
The segmentation/object extraction stage aims at recovering a geometric structure in the image plane that accounts for the desired image characteristics while being in the family of shapes generated by the model . To this end, a mechanism for recovering the most probable object location in the image was considered. Then, one can iterate and move closer to the target by updating the position of the model such that is closer to the desired image characteristics.
The definition of an appropriate image term and the ability to account with various initial conditions are open issues in active shape models. Quite often, their application is based on seeking correspondences between the model and the salient features in the image, a tedious task. More than that, while these models can account in a global manner for the geometric deformations of the object, they fail to capture important local variations.
Level set representations [14] is an established technique for tracking moving interfaces in imaging, vision and graphics [13]. The propagation of curves became popular in computer vision in the late eighties through the introduction of the snake model. Object detection, segmentation and tracking were solved in the context of contour propagation under the influence of internal and external image forces. Level set method is an implicit, intrinsic and parameter free technique to tackle these applications.
One can see numerous advantages for considering a level set variant of the active shape model. Such a formulation could account for various forms (boundary or

regional) of data/image terms of various nature (edges, intensity properties, texture, motion, etc.), an important limitation of the active shape model. Furthermore, one can maintain the implicit and intrinsic property of the level set method as well as the ability to account for topological changes while being able to introduce prior shape knowledge, a task partially addressed up to now [11, 21, 2, 19]. The use of prior knowledge is vital when dealing with corrupted, incomplete or occluded data.
In this paper we propose a level set variant of active shape models. Such a framework is defined on the level set space and consists of various terms. Quite critical is the term that refers to the prior knowledge with objective to constrain the evolving curve to belong to a compact family of shapes - the one recovered through the training set-. Such a term couples two unknown variables; (i) the evolving contour, (ii) the optimal projection parameters of this contour to the model space and imposes the active shape model behavior on the process. Furthermore, various image-driven terms - a major advantage/characteristics of the method - could be considered to guide the evolving contour towards the desired image characteristics.
The most closely related work with our approach, the active shape mode can be found in [5]. In [11, 21, 2, 19] substantial efforts to integrate prior knowledge within level set representations were considered. Worth mentioning is [11, 21] where modeling of prior knowledge is done in a consistent active shape model manner as well as [2, 19] where interesting mathematical formulations introduced to impose prior knowledge. Some comparisons from theoretical as well as experimental point of view will be presented in the discussion section.
The reminder of the paper is organized as follows: in Section 2 we briefly introduce the active shape model and the level set representations, while in Section 3 we address the construction of the prior model in the space of level set functions. The main contribution of the paper, the level set variant of the active shape model is presented in Section 4, while in Section 5 we demonstrate the efficiency and the flexibility of our approach through the integration with various data terms. Discussion appears in Section 6.

2 Overview of Active Shapes & Level Set Methods

2.1 Active Shape Models

Active shape models is a popular object extraction/segmentation technique. One can decompose the technique in two steps, a learning stage and a segmentation stage. Let us assume that a set of training examples that consists of n registered shapes is available $s_{i=1\dots n}$ and let $\overline{s}_{i=1\dots n}$ be a column vector representation of the previous n registered examples according to a sampling rule. Principle Component Analysis (PCA) can be applied to capture the statistics of the corresponding elements across the training examples. PCA refers to a linear transformation of variables that retains - for a given number m of operators - the largest amount of variation within the training data, according to:

$$s(\mathbf{x}) = \overline{s}(\mathbf{x}) + \sum_{j=1}^{m} \lambda_j \ (u_j, v_j) = \overline{s}(\mathbf{x}) + \sum_{j=1}^{m} \lambda_j \ \mathbf{U}_j$$

where $\overline{s}$ is the mean shape, m is the number of retained modes of variation, U_q are these modes (eigenvectors), and λ_j are linear weight factors within the allowable range defined by the eigenvalues. Details on how to recover such a model can be found at [5].

In the active shape model literature one can recover these parameters through an incremental update of the transformation.), $\mathbf{c}_k$ If we assume that for each element of the model $\mathbf{x}_k$, one can recover the corresponding point to the image $\mathbf{c}_k$, then solving the object extraction problem is equivalent first with recovering a similarity transformation $\mathcal{T}$ such that

$$E_{data}(\mathcal{T}) = \sum_{k=1}^{l} |\mathbf{c}_k - \mathcal{T}(\mathbf{x}_k)|^2$$

reaches is lowest potential, where l is the number of basic elements of the model. The most important step in these process is the establishment of correspondences between the actual projection and the true position of the object. Within the original model, one can seek in the normal direction of actual position to recover a most prominent position for this particular model location. Such search is based on the visual profiles, that could been recovered during the model construction.

The Euler-Lagrange equations with respect to the transformation $\mathcal{T}$ parameters $\mathcal{T}_{mn}$, $\frac{\partial}{\partial \mathcal{T}_{mn}} E(\mathcal{T}) = 0$ can provide a close form solution for the estimation of $\mathcal{T}$. To cope for outliers one can consider a robust norm.

Upon convergence of the similarity transformation, one then refines object extraction by seeking a set of coefficients for the principal modes of variation that will move the solution closer to the data. Such an objective is reached through the following cost function,

$$E_{data}(\lambda_1, \lambda_2, ..., \lambda_m) = \sum_{k=1}^{l} \left| \mathbf{c}_k - \mathcal{T}(\mathbf{x}_k) - \sum_{j=1}^{m} \lambda_j \ \mathbf{U}_j \right|^2$$

where the Euler-Lagrange equations $\frac{\partial}{\partial \lambda_j} E_{data}(\lambda_1, \lambda_2, ..., \lambda_m) = 0$ can provide a close form solution for the estimation of λ's.
Recovering correspondences between the model and the image is the most challenging part of such an approach, in particular when the initial conditions are far from the optimal solution. On the other hand, such an approach exhibits robustness, can deal with incomplete and occluded data and is of limited complexity. The solution is equivalent with recovering the global transformation and the coefficients of the modes of variation. We consider a curve propagation variant of this approach through the level set method. We will show that such a method can account for various image/external terms, important shape variations, as well as difficult initial conditions.

2.2 Level Set Representations

Level set representations [14] are a useful mathematical formulation for implementing efficiently curve/surface propagation. Let $[\mathcal{C} : [0,1] \to \mathcal{R}^2,\ p \to \mathcal{C}(p)]$ and let $\mathcal{C}(p,t)$ the state of $\mathcal{C}(p,t)$ driven by the propagation of an initial curve $\mathcal{C}_0(p)$ according to:

$$\mathcal{C}_t(p) \ = \ \mathcal{F}(p)\ \mathcal{N}(p), \quad \mathcal{C}(p,0) \ = \ \mathcal{C}_0(p)$$

where $\mathcal{F}$ is a scalar function and $\mathcal{N}$ the inward normal. One can consider to represent $\mathcal{C}(p)$ with the zero-level set $(\phi \ = \ 0)$ function [14] of a surface z, $\left[z \ = \ (x, y, \phi(x,y,t)) \in \mathcal{R}^3\right]$. Deriving $\phi(x,y,t) \ = \ 0$ with respect to time and space we obtain the following motion for the embedding surface $\phi()$:

$$\phi_t(p) \ = \ \mathcal{F}(p)\ |\nabla \phi(p)|, \quad \phi\left(C_0(p), 0\right) \ = \ 0$$

where $[|\nabla \phi|]$ is the norm of gradient. Such a formulation encodes numerous advantages: (i) the embedding surface $\phi(p)$ remains a function and the evolving contour $\mathcal{C}(p)$ can change topology, (ii) numerical simulations on $\phi(p)$ may be developed and intrinsic geometric properties of the evolving contour can be derived $\phi(p)$ and (iii) the method can be easily extended to deal with problems in higher dimensions.
Furthermore, one can consider the level set space as an optimization framework [22]. Let $\phi : \Omega \times \mathcal{R}^+ \to \mathcal{R}^+$ be a Lipschitz function with the following properties,

$$\phi((x,y);t) = \begin{cases} 0 & ,(x,y) \in \mathcal{C}(t) \\ +\mathcal{D}((x,y), \mathcal{C}(t)) > 0 & ,(x,y) \in \mathcal{C}_{in}(t) \\ -\mathcal{D}((x,y), \mathcal{C}(t)) < 0 & ,(x,y) \in \mathcal{C}_{out}(t) = [\Omega - \mathcal{C}_{in}(t)] \end{cases}$$

where $(x, y) = p$, $\mathcal{C}_{in}(t)$ is the area enclosed by the curve $\mathcal{C}$, $\mathcal{D}((x,y), \mathcal{C}(t))$ the minimum Euclidean distance between the pixel (x, y) and $\mathcal{C}(t)$ at time t. Let us also introduce the approximations of Dirac and Heaviside distributions as defined in [19]. Then one can define terms along $\mathcal{C}$ as well as interior and exterior to the curve using the following properties of the Dirac and Heaviside functions

$$(x, y) \in \Omega : \{\lim_{\alpha \to 0^+} [\delta_\alpha(\phi(x, y)] = 1\} = \mathcal{C}$$
$$(x, y) \in \Omega : \{\lim_{\alpha \to 0^+} [H_\alpha(\phi(x, y))] = 1\} = \mathcal{C}_{in}$$

Such terms will be used later to introduce the active shape drive prior term as well as data/image-driven terms that guides the contour ($\mathcal{C}$) towards the object of interest.

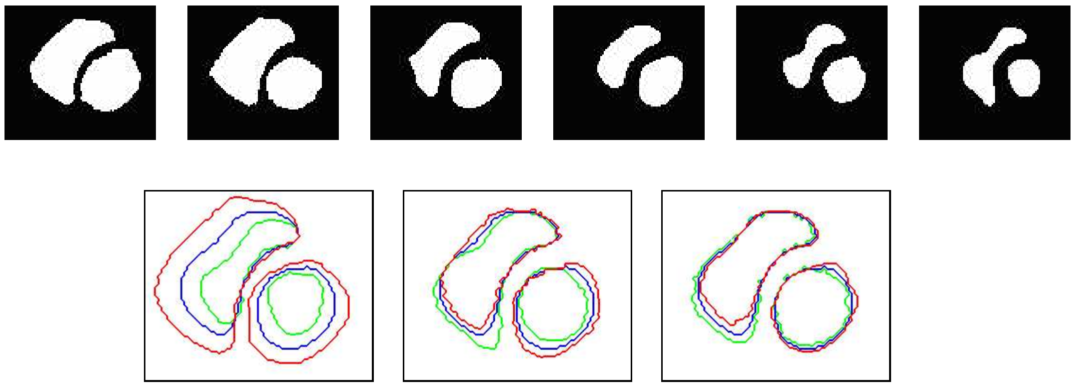
3 Modeling Prior Knowledge in the Level Set Space

Learning the distribution of geometric/image structures is a common problem in computer vision with application to segmentation, tracking, recognition, etc. It is clear that the selection of the representation is important. Given the selected optimization framework, level set functions is a natural selection to account for prior knowledge with numerous earlier described advantages. Let us consider a training set $\mathcal{C}_i$ of N registered curves or surfaces. Then, a distance transform can be used to represent them in the from of a level set function ϕ_i. We consider the method proposed in [17] to recover a rigid alignment between the training contours.
The next step is the construction of the shape model, using the aligned contours. In order to create an invariant representation, one should first normalize the training set ϕ_i. Subtraction of the mean (that can be recovered by averaging ϕ_i's) is a common selection to this end. It is clear that the average level set will not respect the form of the training set, being a distance function. To overcome this limitation, we consider a more vigorous approach [19], seeking to estimate the distance function (ϕ_M) that minimizes:

$$E(\phi_{\mathcal{M}}) = \sum_{i=1}^{n} \iint_\Omega (\phi_i - \phi_{\mathcal{M}})^2 d\Omega, \quad \text{SUBJECT TO} : |\nabla \phi_{\mathcal{M}}|^2 = 1$$

One can optimize such a term though a gradient descent method:

$$\frac{d}{dt}\phi_{\mathcal{M}} = \sum_{i=1}^{n} (\phi_i - \phi_{\mathcal{M}})$$



77% 12% 6%

Figure 1: TOP: Some contours of the Training Set (segmented left ventricle over a cardiac cycle), BOTTOM: Model with the most important shape of variations [principal three modes after rigid alignment (blue:mean, red: $+\sigma$, green:$-\sigma$)] and their contributions.

while its projection to the space of distance functions is done using the following PDE [20]:

$$\left\{ \frac{d}{dt}\phi_{\mathcal{M}} = \left(1 - \text{sign}\left(\phi^0_{\mathcal{M}}\right)\right)\left(1 - |\nabla\phi_{\mathcal{M}}|\right) \right.$$

where $\phi^0_{\mathcal{M}}$ is data driven representation recovered through the gradient descent. The two steps alternate until the system reaches a steady-state solution. Then, we consider the modeling approach introduced in [11, 21]. Once the samples ϕ_i centered with respect to $\phi_{\mathcal{M}}$, $[\psi_i = \phi_i - \phi_{\mathcal{M}}]$, the most important modes of variations can be recovered through Principal Component Analysis:

$$\phi = \phi_{\mathcal{M}} + \sum_{j=1}^{m} \lambda_j \, U_j$$

where m is the number of retained modes of variation, U_j are these modes (eigenvectors), and λ_j are linear weight factors within the allowable range defined by the eigenvalues. An example of this analysis is shown at [figure (1)].

4 Introducing Prior Knowledge in the Level Set Space

Let us now consider an evolving interface represented by a level set function $\phi(x,y)$ as described in Section 2. We would like to evolve it while respecting some shape properties $\phi_{\mathcal{M}}(x,y)$ modulo a transformation $\mathcal{A}$ which belongs to a predefined family. Assuming a rigid transformation with only translation component $\mathcal{A} = (\mathcal{T}_x, \mathcal{T}_y)$ the evolving interface and the transformation should satisfy the following conditions:

$$\begin{cases} (x,y) \to \mathcal{A}(x,y) \\ \phi(x,y) \approx \phi_{\mathcal{M}}(\mathcal{A}(x,y)), \quad \forall (x,y) \in \Omega \end{cases}$$

In that case, the optimal transformation $\mathcal{A}$ should minimize the following functional:

$$E(\phi, \mathcal{A}) = \iint_{\Omega} \rho(\phi, \phi_{\mathcal{M}}(\mathcal{A})) d\Omega$$

where ρ is a dissimilarity measure. A simple choice for ρ is the sum of squared differences. Alternative robust metrics were studied in [15] but without loss of generality we use the sum of squared differences for the sake of simplicity. Scale variation can be added to the rigid transformation $\mathcal{A}$, leading to a similarity one $\mathcal{A} = (\mathcal{S}, \theta, \mathcal{T}_x, \mathcal{T}_y)$. In that case, the objective function should be slightly modified (refer to [19] for further details):

$$E(\phi, \mathcal{A}) = \iint_{\Omega} \left(\mathcal{S}\phi - \phi_{\mathcal{M}}(\mathcal{A})\right)^2 d\Omega$$

Instead of considering the prior on the whole image domain, one can assume that estimating and imposing the prior within the vicinity of the zero-crossing of the level set representation is more meaningful. Within distance transforms, shape information is better captured when close to the origin of the transfromation. The static prior can be thus rewritten:

$$E(\phi, \mathcal{A}) = \iint_{\Omega} \delta_{\epsilon}(\phi) \left(\mathcal{S}\phi - \phi_{\mathcal{M}}(\mathcal{A})\right)^2 d\Omega$$

where $\epsilon \gg \alpha$. The reader will notice that we recover the previous criteria for $\epsilon \to \infty$. In order to minimize the above functional with respect to the evolving level set representation and the global linear transformation, we use the calculus of variations. The current representation will evolve towards $\phi_{\mathcal{M}}$ modulo the rigid transformation $\mathcal{A}$. The equation of evolution for ϕ is given by the calculus of its variations:

$$\frac{d}{dt}\phi = -2\delta_{\epsilon}(\phi)\mathcal{S}\big(\mathcal{S}\phi - \phi_{\mathcal{M}}(\mathcal{A})\big) - \frac{d}{d\phi}\delta_{\epsilon}(\phi)\big(\mathcal{S}\phi - \phi_{\mathcal{M}}(\mathcal{A})\big)^2$$

As mentioned in [19], the second term is a deflation force, and since it does not help in imposing the shape prior, it can be neglected.
The rigid transformation $\mathcal{A}$ is also dynamically updated so as to map ϕ and $\phi_{\mathcal{M}}$ the best. The calculus of variations for the parameters of $\mathcal{A}$ drives to the system:

$$\begin{cases} \frac{d}{dt}\mathcal{S} = 2\iint_\Omega \delta_\epsilon(\phi)(\mathcal{S}\phi - \phi_{\mathcal{M}}(\mathcal{A}))(-\phi + \nabla\phi_{\mathcal{M}}(\mathcal{A}) \cdot \frac{\partial}{\partial \mathcal{S}}\mathcal{A})d\Omega \\ \frac{d}{dt}\theta = 2\iint_\Omega \delta_\epsilon(\phi)(\mathcal{S}\phi - \phi_{\mathcal{M}}(\mathcal{A}))(\nabla\phi_{\mathcal{M}}(\mathcal{A}) \cdot \frac{\partial}{\partial \theta}\mathcal{A})d\Omega \\ \frac{d}{dt}\begin{bmatrix} \mathcal{T}x \\ \mathcal{T}y \end{bmatrix} = 2\iint_\Omega \delta_\epsilon(\phi)(\mathcal{S}\phi - \phi_{\mathcal{M}}(\mathcal{A}))(\nabla\phi_{\mathcal{M}}(\mathcal{A}) \cdot \frac{\partial}{\partial \begin{bmatrix} \mathcal{T}x \\ \mathcal{T}y \end{bmatrix}}\mathcal{A})d\Omega \end{cases}$$

During the model construction, we have analyzed the principal modes of variation within the training set. Including this information, the ideal transformation will map each value of current representation to the "best" level set representation belonging to the class of the training shapes. As mentioned in Section 3, if a shape representation $\phi_{\mathcal{M}}$ belongs to this class, then it can be derived from the principal modes:

$$\phi_{\mathcal{M}} = \phi_{\mathcal{M}} + \sum_{j=1}^{m} \lambda_j \; U_j$$

Hence, the mapping transformation $\mathcal{A}$ must minimize the new energy:

$$E(\phi, \mathcal{A}, \lambda) = \iint_\Omega \left(\mathcal{S}\phi - \left(\phi_{\mathcal{M}}(\mathcal{A}) + \sum_{j=1}^{m} \lambda_j \; U_j(\mathcal{A})\right)\right)^2 d\Omega$$

The weights λ_j will give the "best" level set representation belonging to the learning class. Assuming the evolving level set representation and the rigid mapping fixed, the optimal λ_j are the ones corresponding to the orthogonal projection of ϕ into the PCA subspace (modulo the rigid transformation $\mathcal{A}$):

$$\lambda_j = \iint_\Omega (\mathcal{S}\phi - \phi_{\mathcal{M}}(\mathcal{A}))U_j(\mathcal{A})d\Omega$$

To better account for prior knowledge, one can limit the evaluation of the above cost function in the vicinity of the zero isophote of the average shape by introducing the DIRAC distribution:

$$E(\phi, \mathcal{A}, \alpha) = \iint_\Omega \delta_\epsilon(\phi) \left(\mathcal{S}\phi - \left(\phi_{\mathcal{M}}(\mathcal{A}) + \sum_{j=1}^{m} \lambda_j \; U_j(\mathcal{A})\right)\right)^2 d\Omega$$

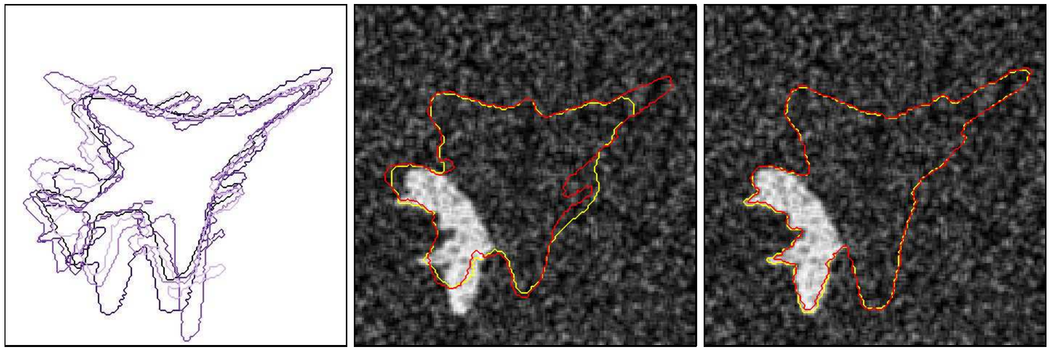


Figure 2: Plane extraction using shape prior and region information: YELLOW: Curve evolution, RED: Projection into the PCA subspace.

Then, the optimal weights λ_j are no more the ones corresponding to the orthogonal projection in the PCA subspace. However, the differentiation gives us a simple linear system:

$$\bar{U}\lambda = b$$

with

$$\begin{cases} \bar{U}(i,j) = \iint_\Omega \delta_\epsilon(\phi) U_i(\mathcal{A}) U_j(\mathcal{A}) \\ b(i) = \iint_\Omega \delta_\epsilon(\phi)(\mathcal{S}\phi - \phi_{\mathcal{M}}(\mathcal{A})) U_i(\mathcal{A}) \end{cases}$$

where $\bar{U}$ is a $m \times m$ positive definite matrix and can be easily inverted. Finally, the minimization of the new energy with respect to the other parameters is identical to the one of the static model.

5 Active Shapes, Level Sets & Object Extraction

In this section, we integrate the proposed level set variant of the active shape model to some of the existing level set segmentation methods that use visual information in various forms [12, 1, 9, 16, 18]. The Geodesic Active Contour model [1, 9] is based on salient features, the Geodesic Active Region model [16] that on top of salient features uses global region statistics, and an approach based on a non-linear feature space [18] which permits to extract textured objects.

(1)

(2)

(3)

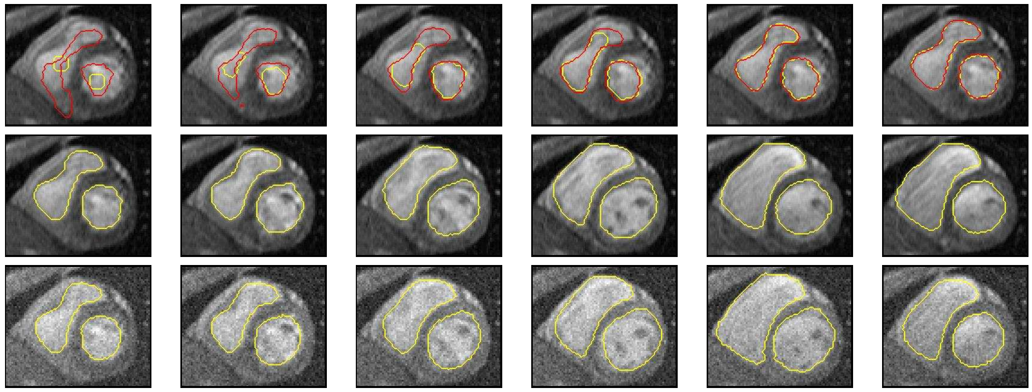


Figure 3: Cardiac tracking: (1) Curve evolution [red] in the first frame and projection to the model space [yellow], (2) Segmentation results over the entire cardiac cycle, (3) Segmentation results over the entire cardiac cycle for corrupted (noisy) data.

5.1 Geodesic Active Contour

Such a framework [1, 9] seeks for a minimal length geodesic curve that is attracted by the desired image properties:

$$E(\phi) = \iint_{\Omega} \delta_{\alpha}(\phi) g(|\nabla I|) |\nabla \phi| d\Omega$$

where g is a decreasing function. One can integrate this term with the prior term introduced in the previous section. While the shape constraint will not change, the PDEs that guide the propagation of the contour will be affected through an extra data-driven/smoothness force.

5.2 Geodesic Active Region

Introducing global regional properties is a common technique to improve segmentation performance. To this end, one can assume a two-class partition problem where the object and the background follow different intensity distribution. Let $p_{\mathcal{C}_{in}}$ and $p_{\Omega-\mathcal{C}_{in}}$ be the densities of $I(x, y)$ in $\mathcal{C}_{in}$ and $\Omega - \mathcal{C}_{in}$. Then according to the Geodesic Active Region model [16] one can recover the object through the optimization of the following function:

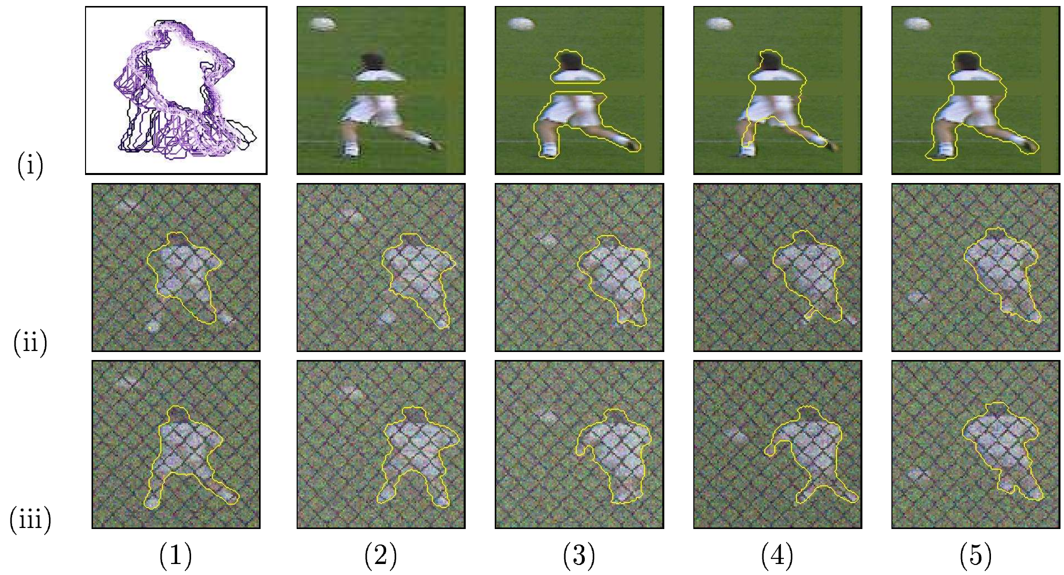


Figure 4: Soccer sequence: [i.1] training set, [i.2] input image, [i.3] geodesic active contour, [i.4] geodesic active contour with the proposed prior, [i.5] geodesic active region with the proposed prior, [ii] segmentation over time presented in a raster scan format using the geodesic active region and the stochastic prior introduced in [19] for corrupted data, [iii] segmentation over time presented in a raster scan format using the geodesic active region and the proposed framework (with 4 modes of variation as shape prior).

$$E(\phi, p_{\mathcal{C}_{in}}, p_{\Omega-\mathcal{C}_{in}}) = (1-a)\iint_{\Omega} \delta_\alpha(\phi) g(|\nabla I|)|\nabla\phi| d\Omega$$
$$- a\iint_{\Omega} \left[H_\alpha(\phi)\log(p_{\mathcal{C}_{in}}(I)) + (1-H_\alpha(\phi))\log(p_{\Omega-\mathcal{C}_{in}}(I))\right] d\Omega$$

One can consider either parametric approximation [16] or a non-parametric density [10] functions to describe the object/background intensity properties. In both cases the new term will result in a local balloon force that moves the contour in the direction that maximizes the posterior segmentation probability as shown in [16]. An extension to vector-valued images was presented in [13] where parametric region densities are estimated adaptively. The image partition is obtained according to color information or any other feature space.

5.3 Non-Linear Feature Space & Texture Segmentation

Indeed, the intensity distribution may not be enough to discriminate the object from the background. Objects with texture information are an example. In that case, one can consider different features that can capture the texture variation. In [18] an efficient non-linear feature space was considered through the diffusion of the following vector:

$$u = (u_1, u_2, u_3, u_4) = \left(I, \frac{I_x^2}{|\nabla I|}, \frac{I_y^2}{|\nabla I|}, \frac{I_{xy}}{|\nabla I|} \right)$$

One can consider such a multi-dimensional feature space for object extraction through the definition of the following image driven term:

$$\begin{aligned} E(\phi, p_{\mathcal{C}_{in}}, p_{\Omega - \mathcal{C}_{in}}) &= (1-a) \iint_\Omega \delta_\alpha(\phi) |\nabla \phi| d\Omega \\ &- a \iint_\Omega \sum_{i=1}^{4} \left[H_\alpha(\phi) \log(p_{\mathcal{C}_{u_i}}(I)) + (1 - H_\alpha(\phi)) \log(p_{\Omega - \mathcal{C}_{u_i}}(I)) \right] d\Omega \end{aligned}$$

where $p_{\mathcal{C}_{u_i}}$ is the density of the object in the u_i feature space and $p_{\Omega - \mathcal{C}_{u_i}}$ the density of the background. Non-linear statistics are used to approximate the visual space (I), while a Gaussian distribution was proven to be efficient for the diffused space. This new term will result in an adaptive balloon force as earlier explained.

5.4 Object Extraction

Each one of these data terms can be used jointly with the shape prior constraint. This data-specific information will make the contour evolve toward the object of intereset while keeping a global shape that belongs to the prior shape family. For this purpose a variational formulation that incorporates two terms is used:

$$E(\phi, \mathcal{A}, \lambda) = b\, E_{shape}(\phi, \mathcal{A}, \lambda) + (1-b)\, E_{data}(\phi)$$

where E_{shape} is the energy presented in Section 4, and E_{data} is one of the three energy accounting for data information. Experimental results of such an integrated framework are presented with each data term: [figure (4)] show results with prior knowledge and a boundary-driven image term, several examples incorportate prior knowledge and boundary/region-driven [figures (2,3,4)], and finally, constrained textured image segmentation is shown in [figure (5)].

6 Discussion

In this paper we have proposed a level set variant of active shape and appearance models to deal with the object extraction. Our approach exhibits numerous advantages. It can deal with noisy, incomplete and occluded data because of its active shape nature. It is intrinsic, implicit parameter and topology free, a natural property of the level set space. Most of all, we have proposed a general formulation where one can integrate different data terms according to the application context. Examples for object extraction, tracking as well as texture segmentation demonstrate the potential of our method.

The proposed framework presents numerous advantages when compared to the prior literature. Introducing image-driven terms of various nature is the added value to the active shapes [5]. Opposite to [11, 21, 2, 19], the level set variant of the active shapes can model complex prior knowledge and being able to account for it. While in [11, 21] the same formulation was considered during the learning phase, object extraction was approached in a static manner. Only the most important models of variation had contributed to the segmentation with static coefficients. Furthermore, in [11] the prior and data-driven term were considered independently. The method alternates between the component aimed to respect the prior and the one moving the contour to the desired image characteristics. In [2, 19], prior knowledge formulation had limited scope (in particular [2] where an average model was used), and therefore the ability to deal with important shape variations was limited [figure 4].

The nature of the sub-space of plausible solutions is a limitation of the proposed framework. Quite often the projection to this space does not correspond to a level set distance function. To account for this limitation, we currently explore prior modeling directly on the Euclidean space, and then conversion to the implicit space during the object extraction.

References

[1] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic Active Contours. In *IEEE ICCV*, pages 694–699, 1995.

[2] Y. Chen, H. Thiruvenkadam, H. Tagare, F. Huang, and D. Wilson. On the Incorporation of Shape Priors int Geometric Active Contours. In *IEEE VLSM*, pages 145–152, 2001.

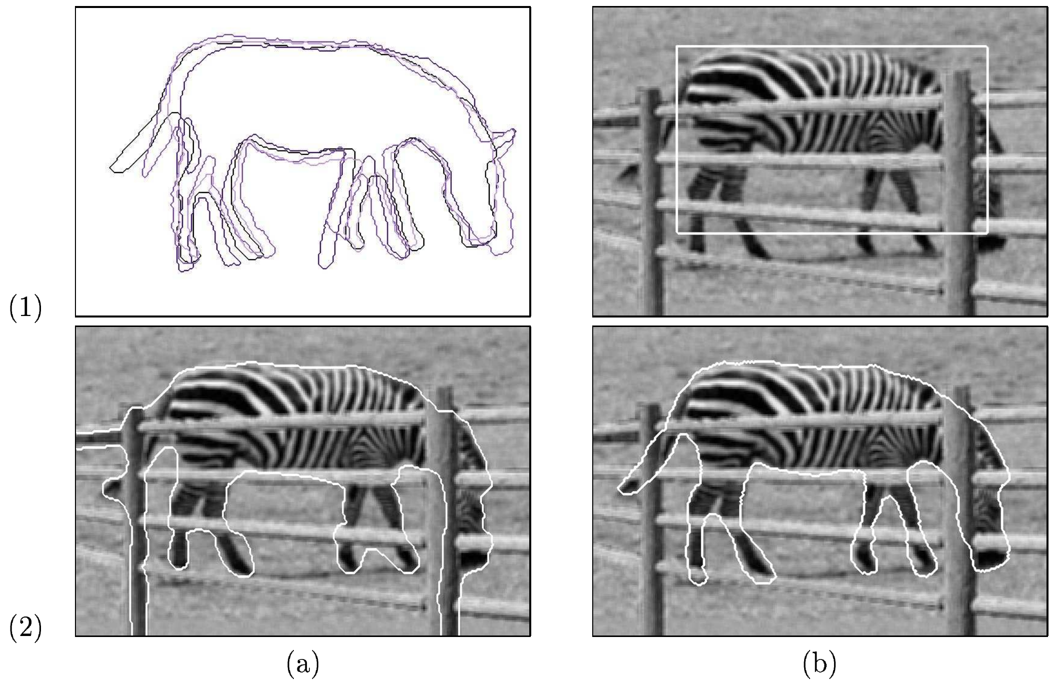


Figure 5: Zebras. [1.a] model, [1.b] initial conditions, [2.a] segmentation [18] without prior knowledge, [2.b] Segmentation [18] using prior knowledge (considering three modes of variation for the shape prior).

[3] T. Cootes, C. Beeston, G. Edwards, and C. Taylor. Unified Framework for Atlas Matching Using Active Appearance Models. In *Information Processing in Medical Imaging*, pages 322–333, 1999.

[4] T. Cootes and C. Taylor. Mixture model for representing shape variation. *Image and Vision Computing*, 17:567–574, 1999.

[5] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61:38–59, 1995.

[6] D. Cremers, T. Kohlberger, and C. Schn *o*rr. Nonlinear Shape Statistics in Mumford-Shah Based Segmentation. In *ECCV*, volume 2, pages 93–108, 2002.

[7] O. Faugeras and R. Keriven. Complete Dense Stereovision Using Level Set Methods. In *ECCV*, pages 379–393, 1998.

[8] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active Contour Models. In *IEEE ICCV*, pages 261–268, 1987.

[9] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi. Gradient flows and geometric active contour models. In *IEEE ICCV*, pages 810–815, 1995.

[10] J. Kim, J. Fisher, A. Yezzi, M. Cetin, and A. Willsky. Non-Parametric Methods for Image Segmentation using Information Theory and Curve Evolution. In *IEEE ICIP*, 2002.

[11] M. Leventon, E. Grimson, and O. Faugeras. Statistical Shape Influence in Geodesic Active Controus. In *IEEE CVPR*, pages I:316–322, 2000.

[12] R. Malladi, J. Sethian, and B. Vemuri. Shape Modeling with Front Propagation: A Level Set Approach. *IEEE PAMI*, 17:158–175, 1995.

[13] S. Osher and N. Paragios. *Geometric Level Set Methods in Imaging, Vision and Graphics*. Springer Verlag, 2003.

[14] S. Osher and J. Sethian. Fronts propagating with curvature-dependent speed : Algorithms based on the Hamilton-Jacobi formulation. *Journal of Computational Physics*, 79:12–49, 1988.

[15] N. Paragios. Shape-based Segmentation and Tracking in Cardiac Image Analysis. *IEEE TMI*, 2003.

[16] N. Paragios and R. Deriche. Geodesic Active Regions: A New Framework to Deal with Frame Partition Problems in Computer Vision. *Journal of Visual Communication and Image Representation*, 13:249–268, 2002.

[17] N. Paragios, M. Rousson, and V. Ramesh. Non-Rigid Registration Using Distance Functions. *Computer Vision and Image Understanding*, 89:142–165, 2003.

[18] M. Rousson, T. Brox, and R. Deriche. Active unsupervised texture segmentation on a diffusion based space. In *IEEE CVPR*, pages II: 699–704, 2003.

[19] M. Rousson and N. Paragios. Shape Priors for Level Set Representations. In *ECCV*, pages II:78–93, Copenhangen, Denmark, 2002.

[20] M. Sussman, P. Smereka, and S. Osher. A Level Set Method for Computing Solutions to Incomprenissible Two-Phase Flow. *Journal of Computational Physics*, 114:146–159, 1994.

[21] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, A. Grimson, and A. Willsky. Model-based Curve Evolution Technique for Image Segmentation. In *IEEE CVPR*, volume I, pages 463–468, 2001.

[22] H.-K. Zhao, T. Chan, B. Merriman, and S. Osher. A variational Level Set Approach to Multiphase Motion. *Journal of Computational Physics*, 127:179–195, 1996.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399